# Scientific Computing III. High Performance Scientific Computing (Phys 2109/Ast 3100H)

### Lecture 2: Parallel Programming Paradigms

SciNet HPC Consortium, University of Toronto

Winter 2013



### Parallel Computers

#### Top500.org:

List of the worlds 500 largest supercomputers. Updated every 6 months,

Info on architecture, etc.



#### Home Lists November 2010

#### TOP500 List - November 2010 (1-100)

R<sub>max</sub> and R<sub>peak</sub> values are in TFlops. For more details about other fields, check the TOP500 description. Power data in KW for entire system

next

-u

Rank	Site	Computer/Year Vendor	Cores	R <sub>max</sub>	Rpeak	Power
1	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C / 2010 NUDT	186368	2566.00	4701.00	4040.0
2	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz / 2009 Cray Inc.	224162	1759.00	2331.00	6950.6
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU / 2010 Dawning	120640	1271.00	2984.30	2580.0
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows / 2010 NEC/HP	73278	1192.00	2287.63	1398.6
5	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz / 2010 Cray Inc.	153408	1054.00	1288.63	2910.0
6	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx super-node S6010/S6030 / 2010 Bull SA	138368	1050.00	1254.55	4590.0

### Supercomputer architectures

- Clusters, or, distributed memory machines
   In essence a bunch of desktops linked together by a network
   ("interconnect"). Easy and cheap.
- Multi-core machines, or, shared memory machines These can see the same memory. Limited number of cores, typically, and much more \$\$\$.
- Vector machines.

These were the early supercomputers, and could do the same operation on a large number of numbers at the same time. Very \$\$\$\$\$, especially at scale.

These days, most chips have some low-level, small size vectorization, but you rarely need to worry about it (compiler should do this).

Most supercomputers are a hybrid combo of these different architectures.



### Distributed Memory: Clusters

# Simplest type of parallel computer to build

- Take existing powerful standalone computers
- And network them



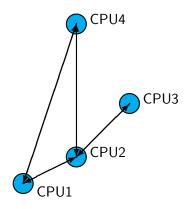
(source: http://flickr.com/photos/eurleif)



#### Distributed Memory: Clusters

#### Each node is independent!

Parallel code consists of programs running on separate computers, communicating with each other. Could be entirely different programs.





### Distributed Memory: Clusters

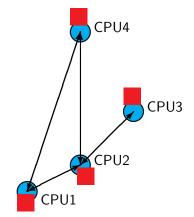
#### Each node is independent!

Parallel code consists of programs running on separate computers, communicating with each other. Could be entirely different programs.

#### Each node has own memory!

Whenever it needs data from another region, requests it from that CPU.

Usual model: "message passing"





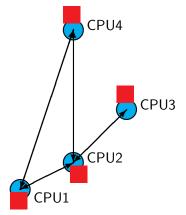
### Clusters+Message Passing

#### Hardware:

Easy to build (Harder to build well) Can build larger and larger clusters relatively easily

#### Software:

Every communication has to be hand-coded: hard to program





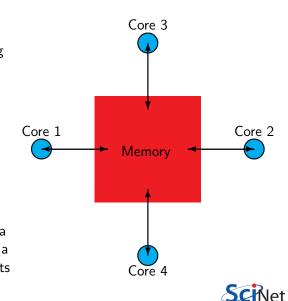
### Shared Memory

One large bank of memory, different computing cores acting on it. All 'see' same data.

Any coordination done through memory

Could use message passing, but no need.

Each code is assigned a thread of execution of a single program that acts on the data.



#### Threads versus Processes

#### Threads:

Threads of execution within one process, with access to the same memory etc.

#### Processes:

Independent tasks with their own memory and resources

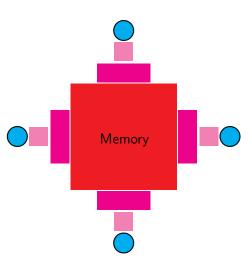
										ljdursl⊚gpc-f102n081:
File	Edit View	Jern	ninal	Tabs	Help					
top -	17:27:34	1 UD 2	da	/s. 1	:40.	1 us	er	. loa	d ave	rage: 1.81. 0.56. 0.20
Tasks	: 142 to:	tal.	3	runnin	a. 13	9 sle	ep	ing.	0 st	opped, 0 zombie
										0.1%hi. 1.0%si. 0.0%st
	1641187									
Swap:		k tot				used.				ee, 2265652k cached
П			ac,			asea,				ce, EEODODEN COCHEG
	USER	PR	NI	VIRT	RES	SHR	s	SCPI1	SMEM	TIME+ COMMAND
	ljdursi			89536						
	root	15		35300					0.0	0:01.57 pbs mom
	root	15		35300				6.0	0.0	0:00,48 pbs mom
	root	15		10344	740	612		0.0	0.0	0:01.45 init
	root	BT	-5	0	0		S	0.0	0.0	0:00.00 migration/0
	root	34	19	0	0		S	0.0	0.0	0:00.00 ksoftirgd/0
	root	BT	-5	0	0		s	0.0	0.0	0:00.00 watchdog/0
	root		-5	6	6		s	0.0	0.0	0:00.01 migration/1
	root	34	19	0	0		s	0.0	0.0	0:00.01 ksoftirgd/1
	root	BT	-5	0	0		S	0.0	0.0	0:00.00 watchdog/1
	root	RT	-5	0	0		S	0.0	0.0	0:00.00 migration/2
	root	34	19	0	0		S	0.0	0.0	0:00.00 ksoftirgd/2
	root	BT	-5	0	0		S	0.0	0.0	0:00.00 ksbrtirdd/2 0:00.00 watchdog/2
							-			
	root		- 5	0		θ	s	0.0	0.0	0:00.00 migration/3
11						0	s			0:00.00 migration/3
11		RT	- 5	θ	8	0	s			0:00.00 migration/3
11 Elle	Edit <u>V</u> iew	RT Term	- 5 iinal	0 Ta <u>b</u> s	0 Help	-		0.0	0.0	0:00.00 migration/3  jdursj@gpc-fl02n081:-
11 Elle   top -	L root Edit ⊻iew 17:33:58	RT Term up 2	-5 inal day	0 Ta <u>b</u> s s, 1:	6 <u>H</u> elp 47,	1 use	er,	0.0	0.0 d aver	0:00.00 migration/3 <u>Jjdursi@gpcsf102n081:</u> age: 0.80, 0.31, 0.17
11 Elle top - Tasks	Edit <u>V</u> iew 17:33:58 : 150 tot	RT Term up 2 al,	-5 inal day 9 r	0 Ta <u>b</u> s s, 1: unning	6 Help 47, , 141	1 use	er,	0.0 loa	0.0 d aver 0 sto	0:00.00 migration/3 Ijdursi@gpc-fl02n081:- age: 0.80, 0.31, 0.17 pped, 0 zombie
11 Elle top - Tasks Cpu(s	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u	RT Term up 2 al, s, θ	-5 linal day 9 r	Ta <u>b</u> s s, 1: unning y, 0.	8 Help 47, , 141 0%ni,	l use L slee	er,	0.0 loar ng, d, 0	0.0 d aver 0 sto .0%wa,	0:00:00 migration/3 Ijdursi@gpc=f102n081;: age: 0:80, 0:31, 0:17 ipped, 0 zombie 0:0%hi, 0:0%si, 0:0%st
Elle top - Tasks Cpu(s Mem:	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872	RT up 2 al, s, 0 k tot	-5 inal day 9 r .0%s al,	Ta <u>b</u> s s, 1: unning y, 0.	8 Help 47, , 141 0%n1, 72k u	1 use l slee used,	er,	0.0 loar ng, d, 0	0.0 d aver 0 sto .0%wa, 0k fro	0:00:00 migration/3 Ijdursi@gpc:f102n081: age: 0.80, 0.31, 0.17 pped, 0 zombie 0.0hii, 0.0%si, 0.0%st re, 256k buffers
11 Elle top - Tasks Cpu(s	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872	RT Term up 2 al, s, θ	-5 inal day 9 r .0%s al,	Ta <u>b</u> s s, 1: unning y, 0.	8 Help 47, , 141 0%n1, 72k u	l use L slee	er,	0.0 loar ng, d, 0	0.0 d aver 0 sto .0%wa, 0k fro	0:00:00 migration/3 Ijdursi@gpc=f102n081;: age: 0:80, 0:31, 0:17 ipped, 0 zombie 0:0%hi, 0:0%si, 0:0%st
Ele ! top - Tasks Cpu(s Mem: Swap:	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872 0	Term up 2 al, s, θ k tot k tot	-5 day 9 r .0%s al, al,	0 Ta <u>b</u> s s, 1: unning y, 0. 28011	6 <u>H</u> elp 47, 1, 141 0%ni, 72k t 0k t	1 use l slee used, used,	er, epi 9%1	0.0 loa ng, d, 0 61070	0.0 d aver 0 stc .0%wa, 0k fre 0k fre	0:00.00 migration/3 [jdurs]_gpc:fl02n0D1: age: 0.80, 0.31, 0.17 pped, 0 zonbie 0.00hi, 0.0%si, 0.0%st e, 2268568k cached
Ele top - Tasks Cpu(s Mem: Swap: Swap:	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872 0 USER	RT Up 2 al, s, 0 k tot k tot	-5 inal day 9 r .0%s al, al,	0 Ta <u>b</u> s s, 1: unning y, 0. 28011 VHRT	6 Help 47, 1, 141 0%ni, 72k t 0k t	1 use L slee used, used, SHR	er, 201 3%1 13	0.0 load ng, d, 0 610700 0	0.0 d aver 0 sto .0%wa, 0% fro 0% fro	0:00.00 migration/3 [jdurs]:gpc:fl02n011: age: 0.80, 0.31, 0.17 pped, 0 zonbie 0.0%hi, 0.0%st, 0.0%st e, 2256 buffers e, 2256 buffers THE E COMMAND
Elle ! top - Tasks Cpu(s Mem: Swap: Swap: 18393	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872 0 USER ljdurs1	RT Jerm up 2 al, s, 0 k tot k tot k tot PR 25	-5 inal day 9 r .0%s al, al, <b>N</b>	0 Ta <u>b</u> s s, 1: unning y, 0. 28011 VIRT 187m	6 Help 47, 1, 14 0%ni, 72k t 0k t 0k t 8 5504	1 use 1 slee 0.0 1sed, 1sed, 1sed, 3484	er, epi 3%1 13 8 R	0.0 load ng, d, 0 6610700 1 80019 100.2	0.0 d aver 0 stc .0%wa, 0% fre 0% fre 0% fre 0.0	0:00.00 nigration/3 igurs15g02710270011: age: 0.80, 0.31, 0.17 popd. 0 200510 0.0Mn1, 0.0Ms1, 0.0Mst e, 256k buffers e, 256k buffers 1716E COMMAD 0:05.45 diffusion-mp1
Elle ! top - Tasks Cpu(s Mem: Swap: Swap: 18393 18395	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872 0 USER 1jdurs1	RT Jerm up 2 al, s, 0 k tot k tot k tot 25 25	-5 inal day 9 r .0%s al, al, 0 0	0 Ta <u>b</u> s s, 1: unning y, 0. 28011 VIRT 187m 187m	6 Help 47, 141 0%n1, 72k t 0k t RES 5504 5512	1 use 0.0 1 stee 1 sed, 1 sed, 1 sed, 3484 3484 3492	er, epi 13 R R	0.0 load ng, d, 0 6610700 0 9CPU 9 100.2 100.2	0.0 d aver 0 stc .0%wa, 0k fre 0k fre 0.0 0.0	0:00.00 migration/3 Idursig:gpc:/102n0013 mge: 0.00, 0.31, 0.17 pped, 0 2005/0 0.0%h1, 0.0%s1, 0.0%st e, 236% buffers e, 236% buffers e
11 Elle   top - Tasks Cpu(s Mem: Swap: PIO 18393 18395 18397	Edit <u>V</u> iew 17:33:58 : 150 tot ):100.0%u 16411872 0 USER ljdurs1 ljdurs1	RT up 2 al, s, θ k tot k tot 25 25 25 25	-5 inal day 9 r .0%s al, al, al, 0 0 0	0 Tabs s, 1: unning y, 0. 28011 28011 187m 187m 187m 187m	6 Help 47, 141 0%n1, 72k t 0k t 8 5504 5512 5508	1 use steed, used, used, 3484 3492 3488	2r, 2pi 3%1 13 R R R R	0.0 load ng, d, 0 6610700 100.2 100.2 100.2	0.0 d aver 0 stc. 0%wa, 0% fre 0% fre 0% 0.0 0.0 0.0	0:00.00 migration/3 jjdursig-opt/102n0115 aps: 0.00, 0.31, 0.17 0.0hh1, 0.0h31, 0.0h31, 0.0h 0.0hh1, 0.0h31, 0.0h 0.0h11, 0.0h31, 0.0h 0.0h14, 0.0h31, 0.0h 0.0h34, 0.0h150.0h 0.0h34, 0.0h150.0h 0.0h150, 0.0h150.0h 0.0h150, 0.0h150.0h 0.0h150, 0.0h150.0h 0.0h150, 0.0h150.0h 0.0h150, 0.0h150,
11 Elle   top - Tasks Cpu(s Mem: Swap: Swap: 18393 18395 18397 18392	Edit View 17:33:58 150 tot ):100.0%u 16411872 0 USER ljdurs1 ljdurs1 ljdurs1 ljdurs1	RT up 2 al, s, θ k tot k tot 25 25 25 25 25	-5 day 9 r .0%s al, al, 0 0 0 0	0 Tabs s, 1: unning y, 0. 28011 28011 187m 187m 187m 187m 187m 187m	e Help 47, 141 0%n1, 72k t 0k t 85504 5512 5508 5580	1 use 1 used, 1 use	2r, 2p1 3%1 13 8 R R R R R R	0.0 load ng, d, 0 6610701 100.2 100.2 100.2 99.9	0.0 d aver 0 stc. 0%wa, 0% fre 0% fre 0.0 0.0 0.0 0.0 0.0	6:08.00 nigration/3 Ijdursig:gp:r/102n0D13 age: 0.080, 0.31, 0.17 pped, 0 2005 0.0%h1, 0.0%st, 0.0%st e, 236% buffers e, 236% buffers e, 236% buffers 0.0%st, 0.0%st 0.0%st, 0.0%st
11 Ele ! top - Tasks Cpu(s Swap: Swap: Swap: 18393 18395 18397 18392 18394	Edit View 17:33:58 : 150 tot ):100.0%u 16411872 0 USER ljdursi ljdursi ljdursi ljdursi ljdursi	RT up 2 al, s, 0 k tot k tot 25 25 25 25 25 25 25	-5 day 9 r .0%s al, al, 0 0 0 0 0	0 Ta <u>b</u> s rs, 1: unning y, 0. 28011 28011 187m 187m 187m 187m 187m 187m	e Help 47, 141 0%n1, 72k t 0k t 85504 5504 5512 5508 5508 5508	1 use 0.0 1 sed, 1 sed, 1 sed, 3484 3492 3488 3556 3488	2r, 2p1 3%1 13 8 R R R R R R R R R	0.0 load ng, d, 0 61070 100.2 100.2 100.2 100.2 99.9 99.9	0.0 d aver 0 stc 0% fre 0% fre 0% fre 0.0 0.0 0.0 0.0 0.0 0.0	0:00.00 migration/3 Jjdursjgpc/102n001: apped. 0.017 opped. 0.017 opped. 0.017 opped. 0.017 opped. 0.017 0.0545 diffusion-mpi 0:05.46 diffusion-mpi 0:05.46 diffusion-mpi 0:05.46 diffusion-mpi
11 Elle 1 top - Tasks Cpu(s Mem: Swap: 18393 18395 18397 18397 18392 18394 18396	Edit View 17:33:58 : 150 tot :100.0%u 16411872 0 USER UJdursi ljdursi ljdursi ljdursi ljdursi	RT up 2 al, s, 0 k tot k tot 25 25 25 25 25 25 25	-5 day 9 r .0%s al, al, 0 0 0 0 0 0	0 Ta <u>b</u> s s, 1: unning y, 0. 28011 187m 187m 187m 187m 187m 187m 187m	Help 47, 1, 143 0%n1, 72k t 0k t 85504 5512 5508 5504 5512 5508 5504 5512	1 use 0.0 1sed, 1sed, 1sed, 3484 3492 3488 3556 3488 3556 3488	2r, 2p1 3%1 13 8 R R R R R R R R R R R R	0.0 load ng, d, 0 610700 100.2 100.2 100.2 100.2 99.9 99.9 99.9	0.0 d aver 0 stc 0 % fre 0 % fre 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	0:00.00 migration/3 Jdursig option/10/200011 appe: 0.80, 0.31, 0.17 pped. 0.200010 0.00011, 0.0001, 0.0001 0.0001, 0.0001, 0.0001 0:05.43 diffusion.mpi 0:05.43 diffusion.mpi 0:05.45 diffusion.mpi 0:05.45 diffusion.mpi 0:05.45 diffusion.mpi
111 Elle 1 top - Tasks Cpu(s Mem: Swap: 18393 18395 18395 18395 18392 18394 18396 18398	Edit <u>V</u> lew 17:33:58 150 tot ):100.0%u 16411872 0 USER ljdursi ljdursi ljdursi ljdursi ljdursi ljdursi	RT up 2 al, s, 0 k tot k tot 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, 0 0 0 0 0 0 0 0 0 0 0	0 Ta <u>b</u> s s, 1: unning y, 0. 28011 187m 187m 187m 187m 187m 187m 187m 18	Help 47, 1, 141 0%n1, 72k t 0k t 885 5504 5512 5508 5504 5512 5508 5504 5512 5508	1 use sleed, ised, ised, ised, ised, ised, ised, ised, 3484 3492 3488 3556 3488 3482 3488 3492 34888 34888 34888 34888 34888 34888 34888	2 r , 2 p 1 3% 1 13 8 R R R R R R R R R R R R R R R R R R R	0.0 load ng, d, 0 60070 100.2 100.2 100.2 100.2 99.9 99.9 99.9 99.9	0.0 d aver 0 stc .0%wa, 0k fre 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Jjdursig-grc/102n0013 age: 0.00, 0.31, 0.17 pped, 0 combe 0.004, 0.004, 0.004 0.004, 0.004 0.004, 0.004 0.004, 0.004 0.005, 0.005 0.005, 0.05
111 Elle ! top - Tasks Cpu(s Mem: Swap: 18393 18395 18395 18396 18396 18398 18398 18398	Edit View 17:33:58 150 tot 1:100.0%u 16411872 0 USER Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi	RT up 2 al, s, θ k tot k tot 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, al, 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 Ta <u>b</u> s s, 1: unning y, 0. 28011 187m 187m 187m 187m 187m 187m 187m 18	Help 47, 141 0%n1, 72k t 0k t 5504 5512 5588 5588 5588 5588 5594 5512 5588 5594 5512	1 use sleed, ised, ised, 3484 3492 3488 3556 3488 3492 3488 3492 3480 3492	27, 201 291 391 13 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	0.0 load ng, d, 0 610700 100.2 100.2 100.2 100.2 99.9 99.9 99.9 99.9 99.9	0.0 d aver 0 stc .0%wa, 0% fre 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Jaursigue/102/0011 aper: 0.00, 0.30, 10.07 pped. 0.20020 0.00h1, 0.00k1, 0.0%st e. 226548 buffers 0:05.43 diffusion.mpi 0:05.46 diffusion.mpi 0:05.46 diffusion.mpi 0:05.46 diffusion.mpi 0:05.45 diffusion.mpi 0:05.45 diffusion.mpi 0:05.45 diffusion.mpi 0:05.45 diffusion.mpi 0:05.45 diffusion.mpi
111 Elle ! top - Tasks Cpu(s Mem: Swap: 18393 18395 18395 18396 18396 18396 18399 1	Edt Vew 17:33:58 150 tot 10:100.0% 16411872 0 USER ljdursi ljdursi ljdursi ljdursi ljdursi ljdursi ljdursi ljdursi	RT up 2 al, s, 0 k tot k tot 25 25 25 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, al, 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 Tabs s, 1: unning y, 0. 28011 187m 187m 187m 187m 187m 187m 187m 18	0 Help 47, 141 0%n1, 72k t 0k t 5504 5512 5508 5504 5512 5508 5512 5508 5512 740	1 use 0.0 1 sleed, 1 sed, 1 sed,	Pr, pi 13 R R R R R R R R R R R R R R R R R R	0.0 load ng, d, 0 610701 100.2 100.2 100.2 100.2 99.9 99.9 99.9 99.9 99.9 99.9 99.9 99.9 9.0	0.0 d aver 0 stc 0% fre 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Ijdursig:gpc;r102n0013 age: 0.80, 0.31, 0.17 pped. 0 zombie 0.0%hi, 0.0%ti, 0.0%ti e, 226% byffres e, 226% byffres e, 226% byffres e, 226% byffres e, 226% byffres 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi 0:05.4% diffrusion-mpi
111 Elle 1 top - Tasks Cpu(s Mem: Swap: Swap: 18395 18397 18395 18397 18396 18398 18398 18398 18398 122 122 123 123 123 123 123 123	Edit View 17:33:58 150 tot ):100.0%w 16411872 0 USER Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi Ujdursi	RT up 2 al, s, 0 k tot k tot PR 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, 0 0 0 0 0 0 0 0 0 0 0 0 0	0 Tabs s, 1: unning y, 0. 28011 187m 187m 187m 187m 187m 187m 187m 18	0 <u>H</u> elp 47, 1,141 0%n1, 72k ( 0k ( 0k ( 5504 5504 5512 5508 5504 5512 5504 5512 5504 65512 65502 65504 65512 65502 65502 65512 65502 655	1 use 0.0 1 steed, 0.0 1 steed, 1 steed,	SRRRRRRRRSS	0.0 load ng, d, 0 6610700 100.2 100.2 99.9 90.9 90.0 80.	0.0 d aver 0 stc 0 hwa, 0 h fro 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Jdursigcpcfl02n0111 ape: 0.80, 0.31, 0.17 ope: 0.80, 0.31, 0.05 officient officient officient officient officient officient officient officient officient officient officient officient 0:05.46 diffusion-mpil 0:05.46 dif
111 Ele 1 top - Tasks Cpu(s Mem: Swap: B100 18393 18395 18395 18397 18392 18394 18396 18398 18399 1 13994 1 2 3 3	Edit View 17:33:58 17:33:59 tot 17:33:50 tot 19:00.0% 10:00.0	RT up 2 al, 0 k tot k tot 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, 0 0 0 0 0 0 0 0 0 0 0 0 0	0 Tabs rs, 1: unning 28011 28011 187m 187m 187m 187m 187m 187m 187m 18	e <u>Help</u> 47, 141 0%n1, 5504 5504 5512 5588 5588 5598 5598 5598 5598 5598 5598 5598 60 5512 5599 0 0 0 0 0 0 0 0 0 0 0 0 0	1 use 0.0 sed, 0.0 sed, 3484 3492 3488 3556 3488 3492 3480 3492 612 0 0	SRRRRRRSSS	0.0 load ng, d, 0. 610700 100.2 100.2 100.2 100.2 99.9 99.9 99.9 99.9 99.9 99.9 99.9 0.0 0.0	0.0 d aver 0 sto 0 maa, 0 k fro 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Jauris:pre:rdo2not1: gen: 0.80, 0.31, 0.17 pped. 0.2003.0.41, 0.0%st e. 2565 buffers 0:05.45 diffusion-mpi 0:05.45 diffusion-mpi 0:05.46 diffusion-mpi
111 Ele   top - 1 Tasks Cpu(s Mem: Swap: 18392 18395 18397 18392 18394 18398 18398 18398 18398 123 23 34	Edit View 17:33:58 150 tot ):100.0%w 16411872 0 USER Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1 Ujdurs1	RT up 2 al, s, 0 k tot k tot 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 Tabs s, 1: unning y, 0. 28011 187m 187m 187m 187m 187m 187m 187m 18	e Help 47, , 141 0%n1, 0%n1, 0%n1, 5504 5512 5508 5512 5508 5512 5509 5512 740 0 0 0 0	1 use l slee l sleed, ised, ised, ised, 3488 3482 3488 3492 3488 3492 3488 3492 612 0 0 0 0 0	SRRRRRRSSSS	0.0 load ng, d, 0 610701 100.2 100.2 99.9 99.9 99.9 99.9 99.9 99.9 99.9 0.0 0.0	0.0 1 aver 0 sto 0/ma 0k fro 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Jdursigor/J02n0D1: age: 0.80, 0.31, 0.17 pped. 0 20010 0.000 0.000 0.000 0.000 0:05.45 diffusion-mpi 0:05.45 diffusion-mpi
111 Elle   Tasks Cpu(s Mem: Swap: 18395 18395 18397 18392 18394 18398 18399 1 18398 18399 1 2 3 3 4 5	Edit View 17:33:58 17:33:59 tot 17:33:50 tot 19:00.0% 10:00.0	RT Jerm i up 2 al, s, 0 k tot k tot 25 25 25 25 25 25 25 25 25 25	-5 inal day 9 r .0%s al, al, 0 0 0 0 0 0 0 0 0 0 0 0 0	0 Tabs rs, 1: unning 28011 28011 187m 187m 187m 187m 187m 187m 187m 18	e <u>Help</u> 47, 141 0%n1, 5504 5504 5512 5588 5588 5598 5598 5598 5598 5598 5598 5598 60 5512 5599 0 0 0 0 0 0 0 0 0 0 0 0 0	1 use l slee l sleed, ised, 3484 3492 3488 3492 3488 3492 3488 3492 3488 3492 612 0 0 0 0 0 0 0	SRRRRRRSSS	0.0 1004 100.2 100.2 99.9 99.9 99.9 99.9 99.9 99.9 0.0 0.0	0.0 d aver 0 sto 0 maa, 0 k fro 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.	0:00.00 migration/3 Jauris:pre:doi:00.00 Jauris:pre:doi:00.00 0:00.00 0:00.00 0:05.43 diffusion-mpi 0:05.45 diffusion-mpi



### Shared Memory: NUMA

#### Non-Uniform Memory Access

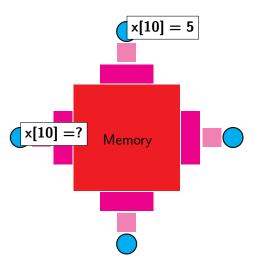
- Each core typically has some memory of its own.
- Cores have cache too.
- Keeping this memory coherent is extremely challenging.





### Coherency

- The different levels of memory imply multiple copies of some regions
- Multiple cores mean can update unpredictably
- Very expensive hardware
- Hard to scale up to lots of processors, very \$\$\$
- Very simple to program!!





### Shared Memory Communication Cost

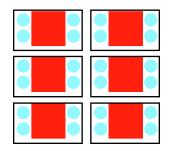
	Latency	Bandwidth
GigE	10 $\mu$ s	1 Gb/s
	(10,000 ns)	( 60 ns/double)
Infiniband	2 µs	2-10 Gb/s
	(2,000 ns)	(10 ns /double)
NUMA	0.1 <i>µ</i> s	10-20 Gb/s
(shared memory)	(100 ns)	(4 ns /double)

Processor speed:  $O(GFLOP) \sim$  few ns or less.



### Hybrid Architectures

- Multicore machines linked together with an interconnect
- Many cores have modest vector capabilities.
- Machines with GPU or other coprocessors: GPU is multi-core, but the amount of shared memory is limited.







 Embassingly parallel applications Scripting, GNU Parallel<sup>1</sup>



 Embassingly parallel applications Scripting, GNU Parallel<sup>1</sup>

 Shared memory OpenMP, Threads, Automated parallelization



 Embassingly parallel applications Scripting, GNU Parallel<sup>1</sup>

 Shared memory OpenMP, Threads, Automated parallelization

 Distributed memory MPI, Files



- Embassingly parallel applications Scripting, GNU Parallel<sup>1</sup>
- Shared memory OpenMP, Threads, Automated parallelization
- Distributed memory MPI, Files
- Graphics computing CUDA, OpenCL



- Embassingly parallel applications Scripting, GNU Parallel<sup>1</sup>
- Shared memory OpenMP, Threads, Automated parallelization
- Distributed memory MPI, Files
- Graphics computing CUDA, OpenCL
- Hybrid combinations

<sup>1</sup>O. Tange (2011): *GNU Parallel* - *The Command-Line Power Tool*, ;login: The USENIX Magazine, February 2011:42-47.



Data or computation bound?

#### Computation bound $\Rightarrow$

#### Task parallelism

Focus on processes/threads. These processes may have quite different computations to do. Bring the data to the computation

#### Data bound $\Rightarrow$

#### Data parallelism

Focus on operations on a given, large data set. Data set often an array, partitioned, and tasks act on separate partitions. Bring the computation to the data



#### Granularity

Fine-grained parallelism

Small individual tasks.

The data is transferred among processors frequently. *Often OpenMP, to overlap different hardware functions* 

Coarse-grained parallelism

Data communicated infrequently, after larger amounts of computation. *Often MPI, because of network latency* 

Too fine-grained  $\Rightarrow$  overhead Too course-grained  $\Rightarrow$  load imbalance

Balance depends on architecture, access pattern, and computation



# Using SciNet





### Using SciNet

#### **GPC**



- 3780 nodes each with 2× 2.53GHz quad-core Intel Xeon 5500 64-bit processors
- 30240 cores total
- 16GB RAM per node
- No local hard disks
- Infiniband network on all nodes
- 306 TFlops
- ▶ #66 on Nov 2012 *TOP500* supercomputer list

Mini intro to SciNet - how to get started

- Need to have an account
- If you don't: get it (wiki.scinethpc.ca/wiki/index.php/Essentials)
- If you can't: email us.
- Read the SciNet Tutorial and the GPC quick start on the wiki. (wiki.scinethpc.ca/wiki/index.php/GPC\_Quickstart)

Access:

```
$ ssh -X login.scinet.utoronto.ca
$ ssh -X gpc01 (or gpc02, gpc03, gpc04)
```



#### Mini intro to SciNet - how to compile

- ► Ssh into one of the devel nodes gpc0{1,2,3,4}.
- Then load a compiler into your environment:

\$ module load gcc

- ► Then you can compile using g++.
- ► In general, the Intel compilers are preferred on the GPC, but for the purpose of this course, g++ is fine too.



#### Mini intro to SciNet - how to run

- You do not run on the devel nodes.
- You run on compute nodes.
- Compute nodes are reserved through the queue.
- You can get an interactive session on a compute nodeby making the following request to the scheduler (from one of the devel nodes)

```
$ qsub -1 nodes=1:ppn=8,walltime=2:00:00 -I -X -qdebug
```

which gets a dedicated compute node for two hours. You can ask for more time but you will need to omit the '-qdebug' and you may be waiting for a long time for a node.

Alternatively, submit a job script.



#### **GNU** Parallel

- What if you need to keep all 8 cores on a node busy.
- GNU Parallel can help you with that!
- GNU parallel is a really nice tool to run multiple serial jobs in parallel. It allows you to keep the processors on each 8core node busy, if you provide enough jobs to do.
- GNU parallel is accessible on the GPC in the module gnu-parallel.

```
$ module load gnu-parallel/20120622
```

Note that we recommend the newer version of gnu-parallel over the (default) 2010 one.



#### **GNU** Parallel Example

```
gpc-f101n084-s g++ -O3 mycode.cc -o mycode
gpc-f101n084-s mkdir sSCRATCH/example2
gpc-f101n084-s cp mycode sSCRATCH/example2
gpc-f101n084-s cd sSCRATCH/example2
gpc-f101n084-s cat > joblist.txt
  mkdir run1; cd run1; ../mycode 1 > out
mkdir run2; cd run2; ../mycode 2 > out
  mkdir run64; cd run64; ../mycode 64 > out
gpc-f101n084-s cat > myjob.pbs
    #!/bin/bash
    #PBS -1 nodes=1:ppn=8,walltime=24:00:00
    #PBS -N GPJob
    cd sPBS_O_WORKDIR
    module load gcc gnu-parallel/20120622
    parallel -j 8 < joblist.txt
gpc-f101n084-s qsub myjob.pbs
    2961985.gpc-sched
gpc-f101n084-s ls
    GP.Job.e2961985
                      GPJob.o2961985 joblist.txt
                      myjob.pbs/
                                       řun3/
    mycode
```



### GPC and hyperthreading

- HyperThreading: Appears as if there are 16 processors rather than 8 per node.
  - For OpenMP applications this is the default unless OMP\_NUM\_THREADS is set.
  - ► For MPI, try -np 16.
  - ▶ For gnu parallel, use -j 16.
- ► Always request ppn=8, even with hyperthreading.
- Always test if this is beneficial and feasible!



- Make sure you've got a SciNet account!
- Read the SciNet tutorial (as it pertains to the GPC)
- Read the GPC Quick Start.
- Get the first set of code:

s cd \$SCRATCH
s git clone /scinet/course/sc3/homework1
s cd homework1
s source setup
s make

- Contains threaded program 'blurppm' and 266 ppm images to be blurred.
- Usage:

blurppm INPUTPPM OUTPUTPPM BLURRADIUS NUMBEROFTHREADS



Simple test:

```
$ qsub -l nodes=1:ppn=8,walltime=2:00:00 -I -X -qdebug
$ cd $SCRATCH/homework1
$ blurppm 001.ppm new001.ppm 30 1
$ display 001.ppm &
$ display new001.ppm &
```

Will want to do timing tests:

```
$ time blurppm 001.ppm new001.ppm 30 1
real 0m52.900s
user 0m52.881s
sys 0m0.008s
```



#### Part 1

- Time blurppm with a BLURRADIUS ranging from 1 to 41 in steps of 4, and for NUMBEROFTHREADS ranging from 1 to 16. Record the (real) duration of each run.
- Plot the duration as a function of NUMBEROFTHREADS, as well as the speed-up and efficiency.
- Submit script and plots of the duration, speedup and effiency as a function of NUMBEROFTHREADS.



Part 2

- Use GNU parallel to run blurppm on all 266 images with a radius of 41.
- Investigate different scenarios:
  - 1. Have GNU parallel run 16 at a time with just 1 thread.
  - 2. Have GNU parallel run 8 at a time with 2 threads.
  - 3. Have GNU parallel run 4 at a time with 4 threads.
  - 4. Have GNU parallel run 2 at a time with 8 threads.
  - 5. Have GNU parallel run 1 at a time with 16 threads.

Record the total time it takes in each of these scenarios.

- Repeat this with a BLURRADIUS of 3.
- Submit scripts, timing data and plots.

